17K Mobile Games

Midterm Project

Yiru Fei

11/5/2019

1. Abstract

Before downloading a mobile game, the players' first concern is usually the rating of games. In another word, a game with a lower rating is less likely to be download than a game with a higher rating. This project will mainly focus on which factor of the first impression may affect whether a player download the mobile games. The first impression only includes information that players can get from the app store. Based on the data from the Apple App Store, this project will analyze with EDA method and modeling method.

2. Introduction

Towards the end of the 20th century, mobile phone ownership became ubiquitous in the industrialized world. As a result of this explosion, technological advancement by handset manufacturers became rapid, and mobile phone games also became increasingly sophisticated, taking advantage of exponential improvements in display, processing, storage, interfaces, network bandwidth, and operating system functionality. The mobile games industry is worth billions of dollars, with companies spending vast amounts of money on the development and marketing of these games to an equally large market.

There is no doubt that a good game must include great controls, interesting themes, excellent music, the balance of challenge, and entertaining story. But when players can only obtain game information from the APP store, how do they determine whether to download the game? Therefore, exploring the relationship between game rating and first impression of games is significant.

3. Method

Data source

This data is obtained from https://www.kaggle.com/tristan581/17k-apple-app-store-strategy-games. It is the data of 17007 strategy games on the Apple App Store, and collected on the 3rd of August 2019, using the iTunes API and the App Store sitemap.

-Name: the name of game

-Average User Rating: Rounded to nearest 0.5, requires at least 5 ratings

-User Rating Count: Number of ratings internationally, null means it is below 5

-Price: Price in USD

-Description: App description

-Developer: App developer

-Languages: ISO2A language codes

-Size: Size of the app in bytes.

-Age Rating: Either 4+, 9+, 12+ or 17+

-Original Release Date: When it was released

-Current Version Release Date: When it was last updated

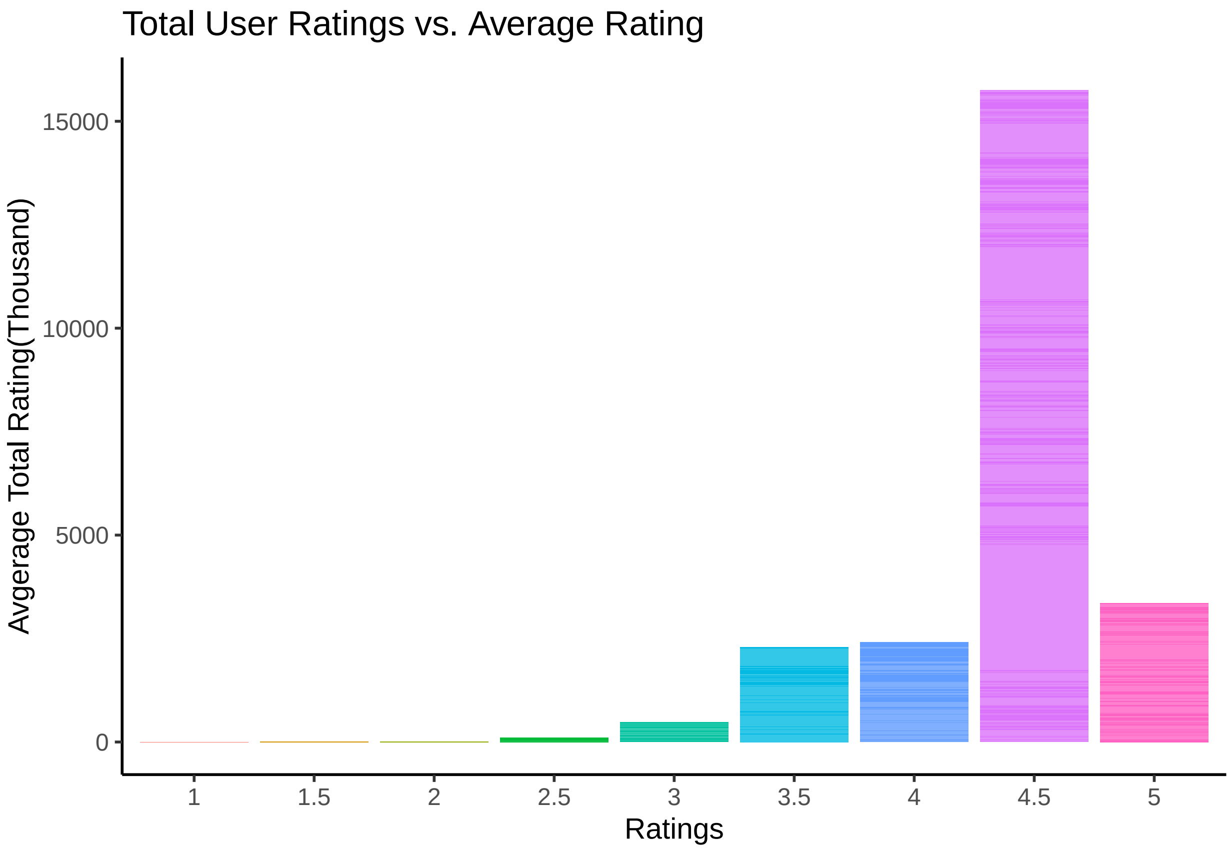
After eliminating the game with 0 rating, and duplicate games, the data set has 7220 observations and 18 variables.

EDA

Fig.1 Average Rating vs. Price

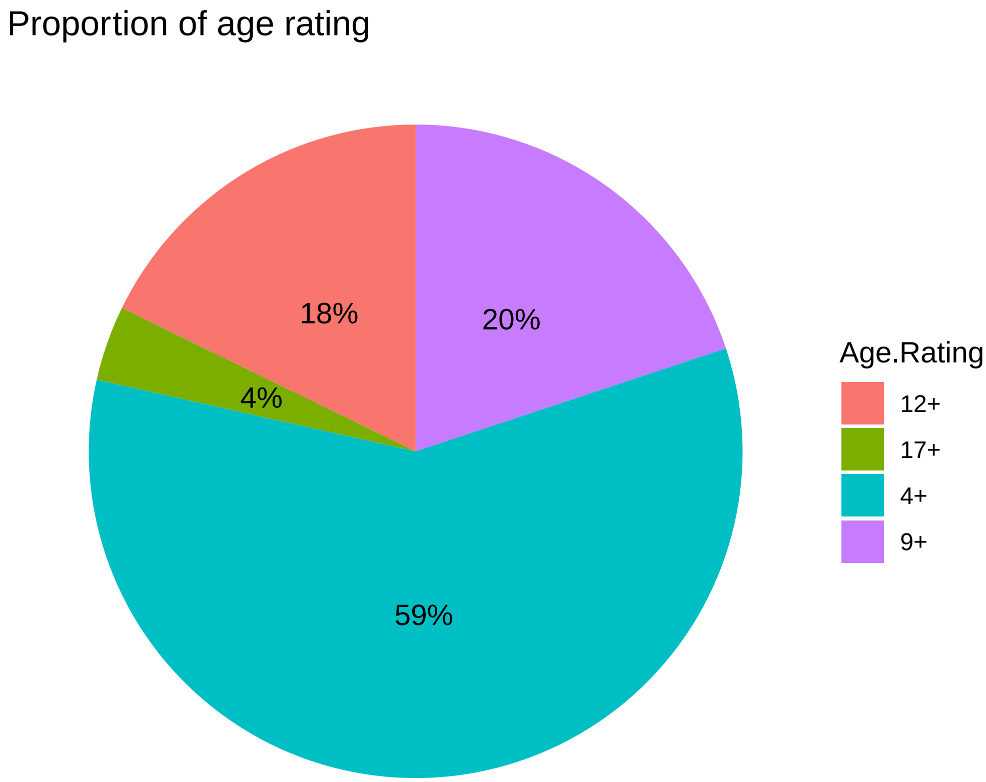
By divided all games into paid games and free games, then build a boxplot between these two groups and the average rating. From the boxplot, the quartile range of two boxplots are the same, and both plots have outliers. However, the medians level in two plots is different. Free games have a higher average rating than paid games.

Fig.2 Total User Rating vs. Average Rating



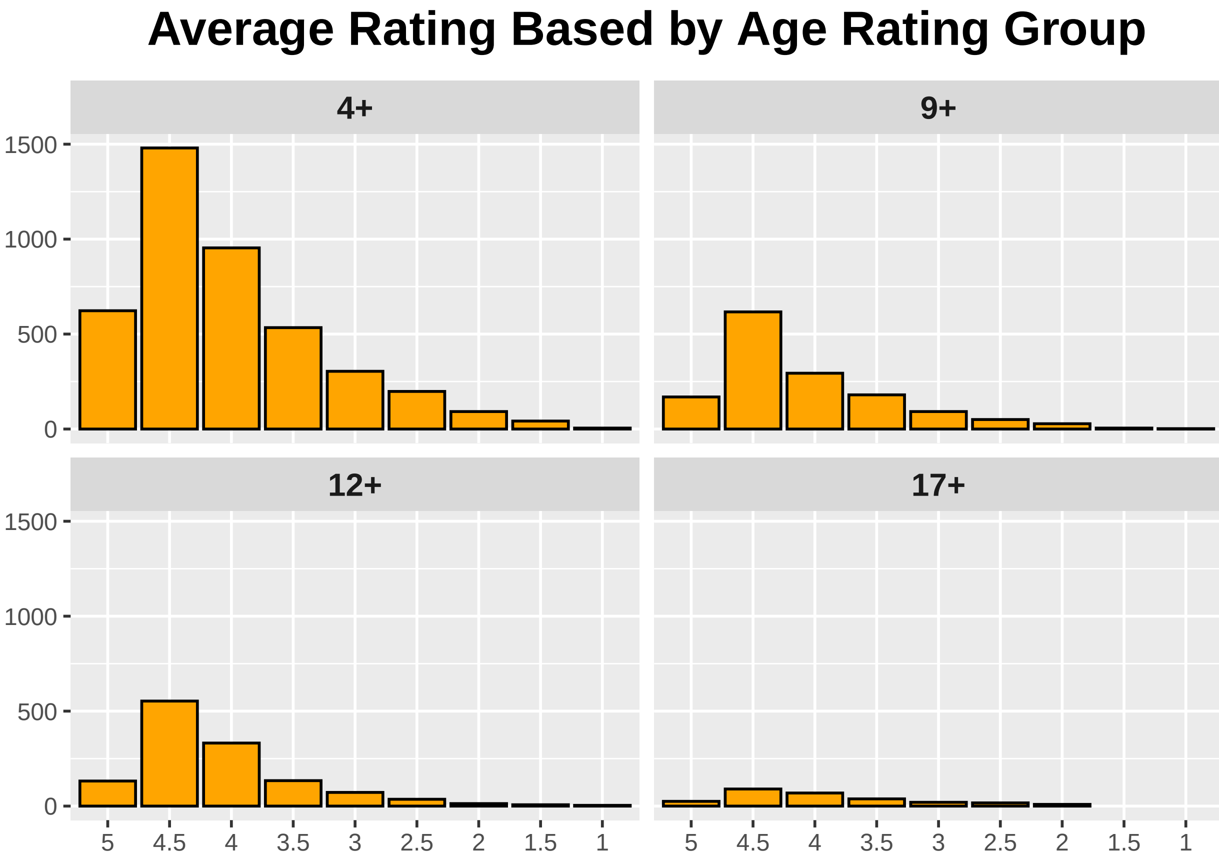
Since this data was collected from 2003 to 2019, there are more than tens of millions of reviews for whole games. I divided the total rating by 1000 to making the number on the y-axis smaller. From figure 2, the most players will rate 4.5 for a game, which total rating is over $1.5 * 10^7$, and the total number of rating on 3.5 and 4 are closed.

Fig.3 Pie Chart of age rating



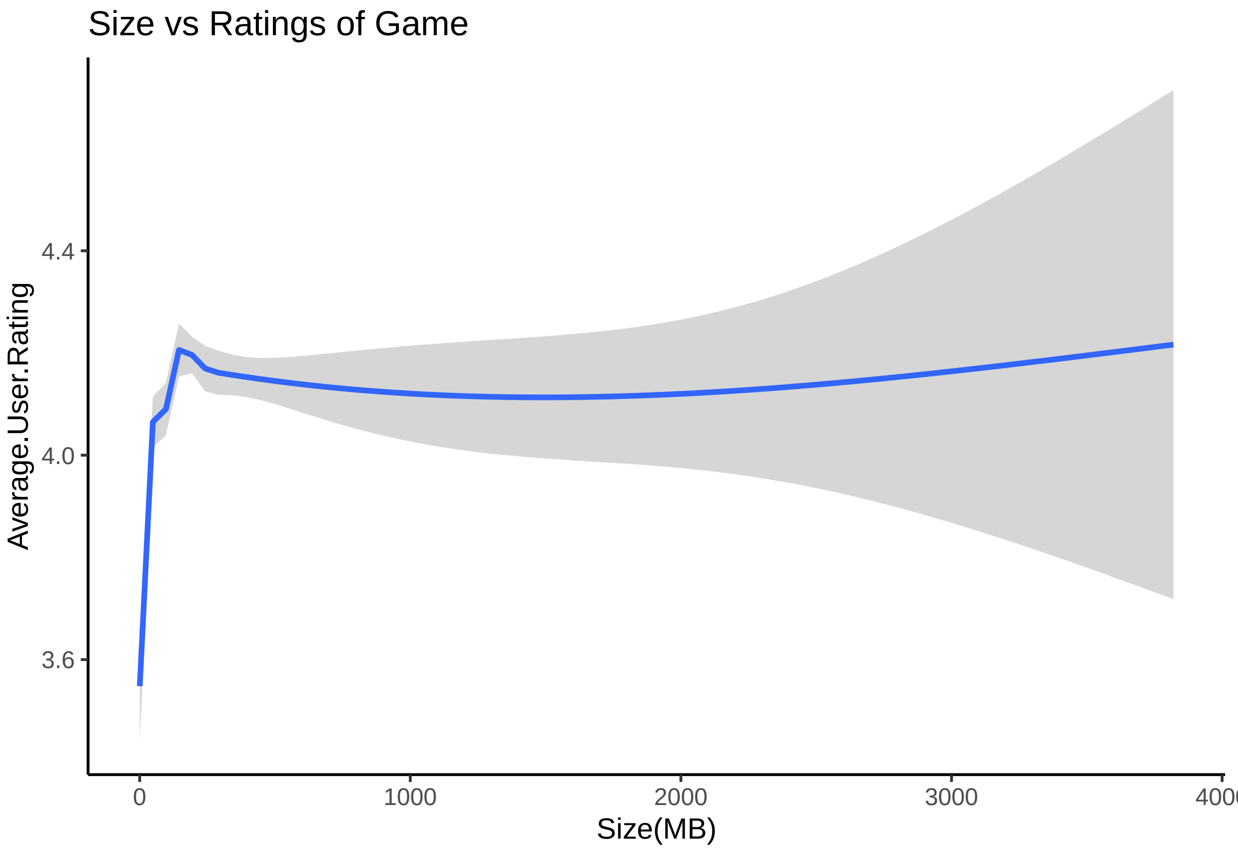
Before checking whether players' age may affect the game rating, first of all, I need to know the proportion of different age ratings in this data. From figure 3, the most minimal age for downloading games are four years old or older.

Fig.4 Average Rating Based by Age Rating Group



After grouping games by age rating, and making barplot between the total number of rating and average rating again, I can see the game rating in different age groups have a similar shape, all groups have the highest number on 4.5. Figure 4 indicates that the age rating will not affect game rating a lot.

Fig.5 Size vs Ratings of APP



In the raw data, the unit of app size is in byte. I divide the total size by $1024 * 1024$ for transforming the unit into MB and then make figure 5. This plot indicates that pretty small size games are hard getting a high rate, and when the game size is larger than 300 MB approximately, their ratings will not have a strong relationship with their size.

Fig.6 Languages available in apps vs. Rating of Game

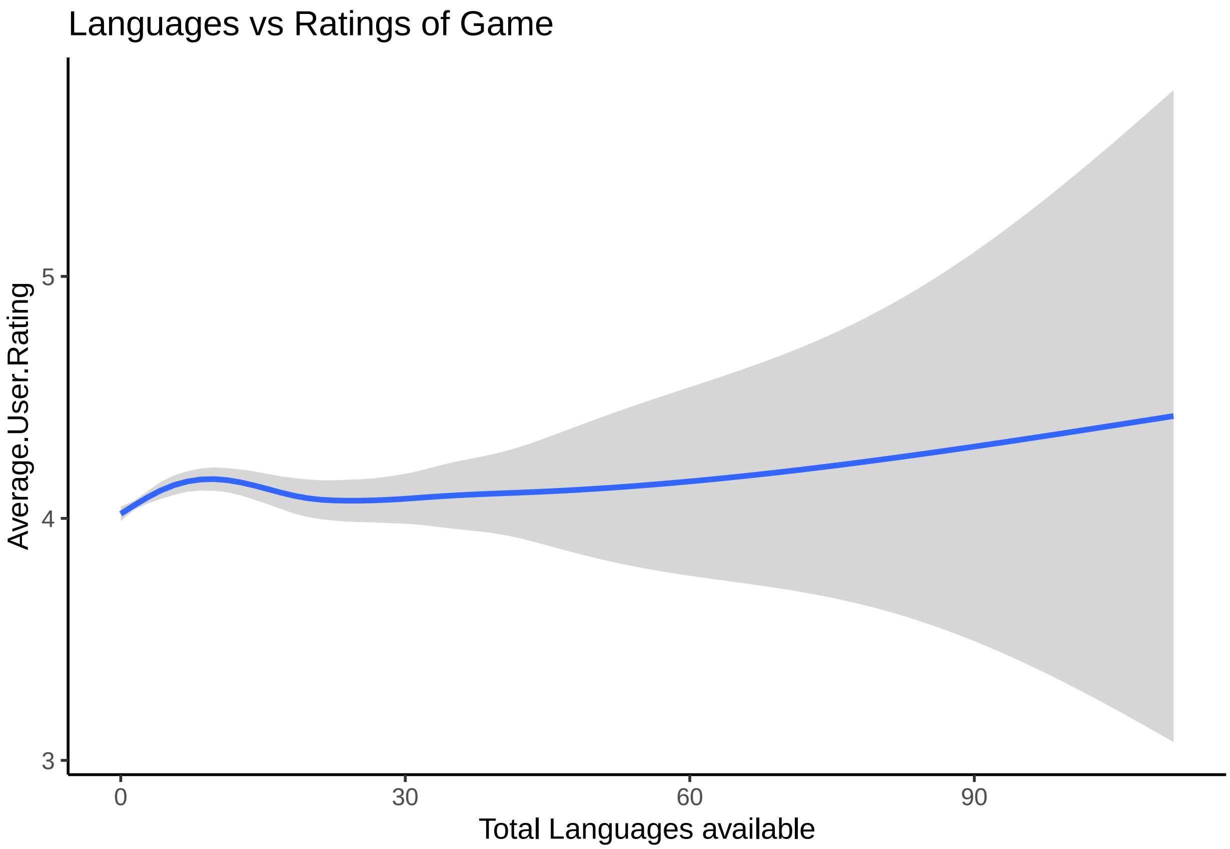


Figure 6 indicates that games include more languages have a wider range of rating than games have fewer languages. Maybe, due to if a game contains more languages, then they will have more players and more ratings. Therefore, those games have different grades.

Fig.7 Total Words on Description vs. Rating of Game

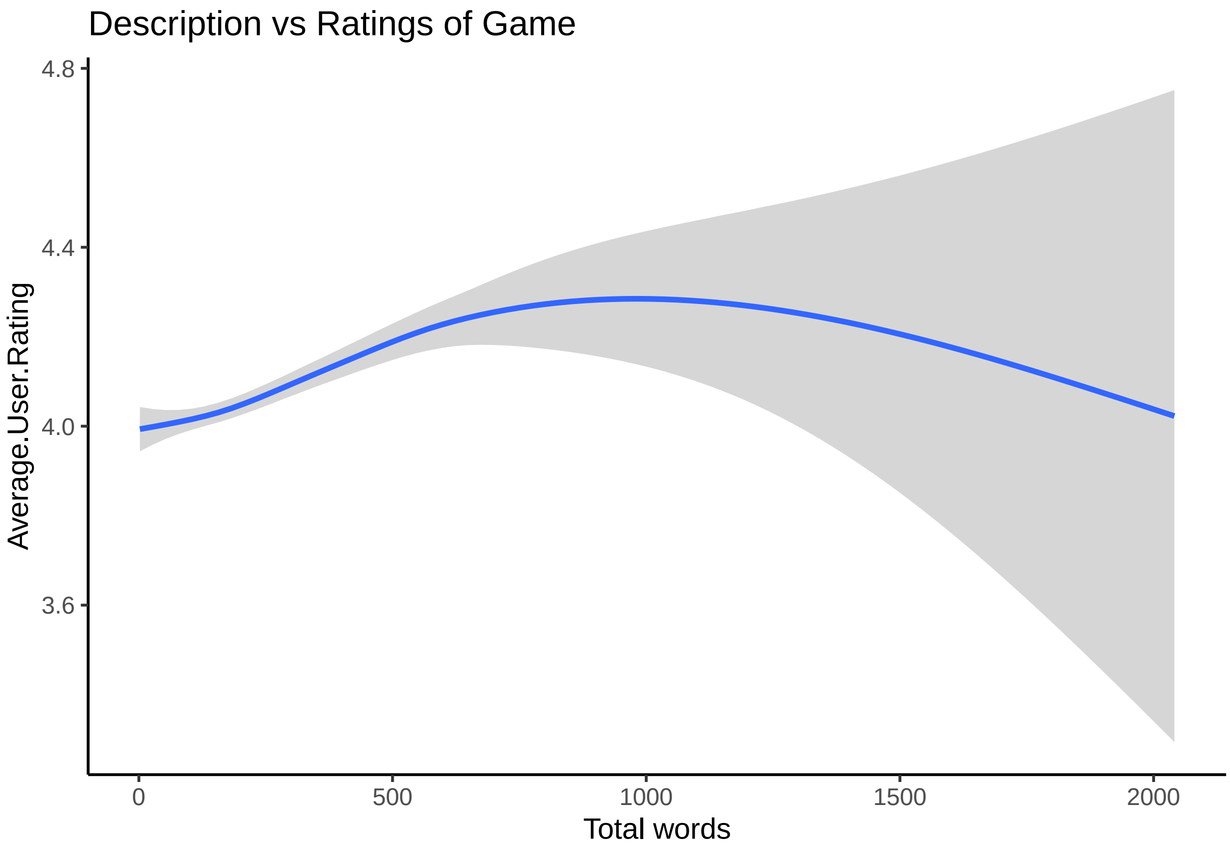


Figure 7 indicates that games have more specific descriptions more substantial possible on getting a lower rating than games have fewer descriptions. In other words, even if a game has only one sentence in summary, it may not get a low rating.

Fig.8 Average Rating vs. Original Release Date

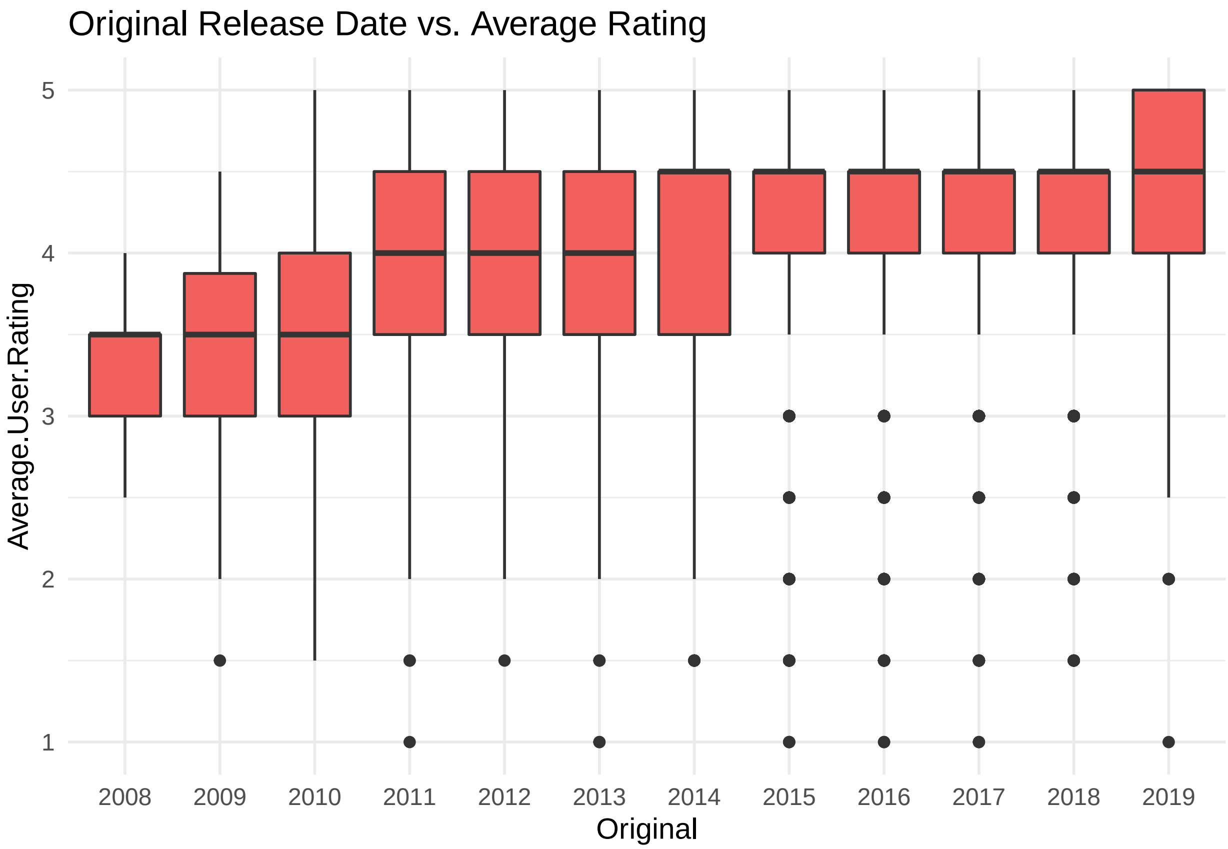


Figure 8 shows that new released games have higher average ratings than old ones.

Fog.9 Average Rating vs. Current Release Date

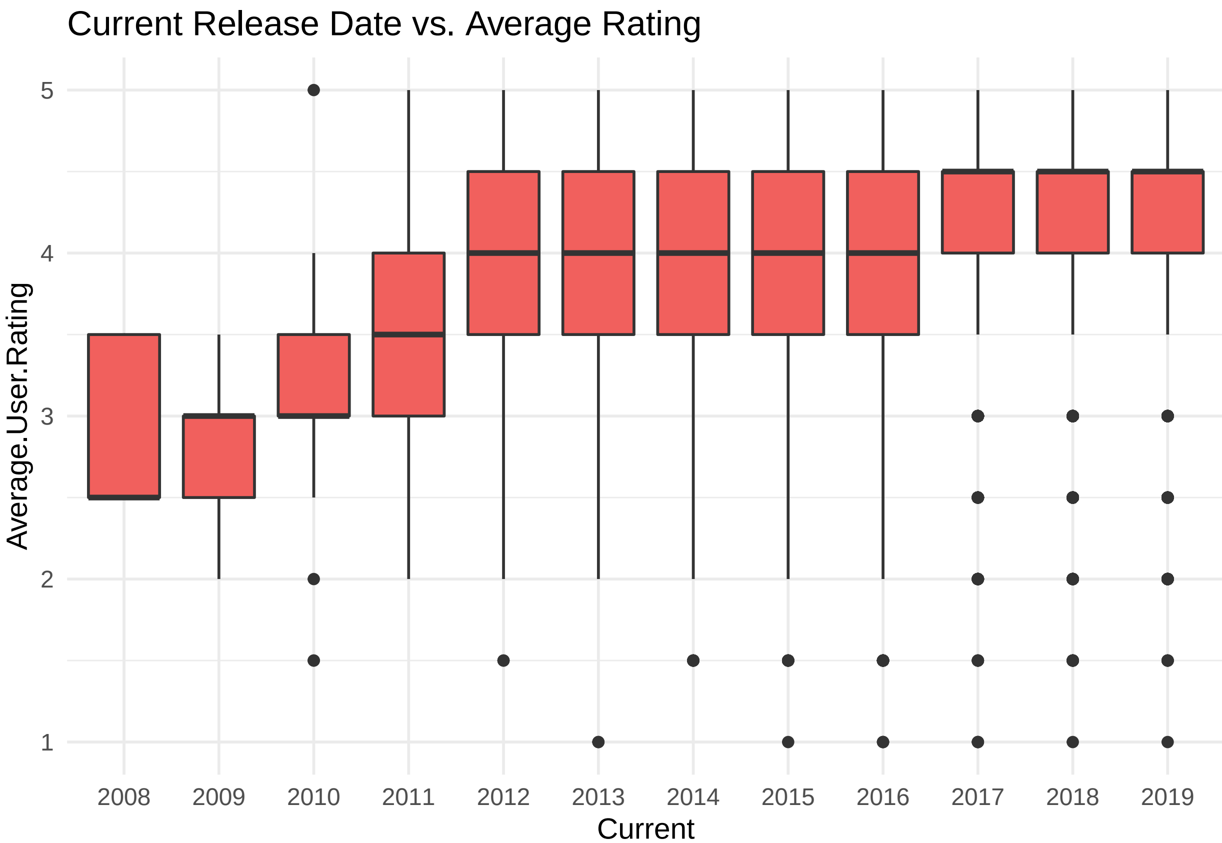


Figure 9 shows a similar result as figure 8, games that have been updated will have higher ratings than games that have stopped updating.

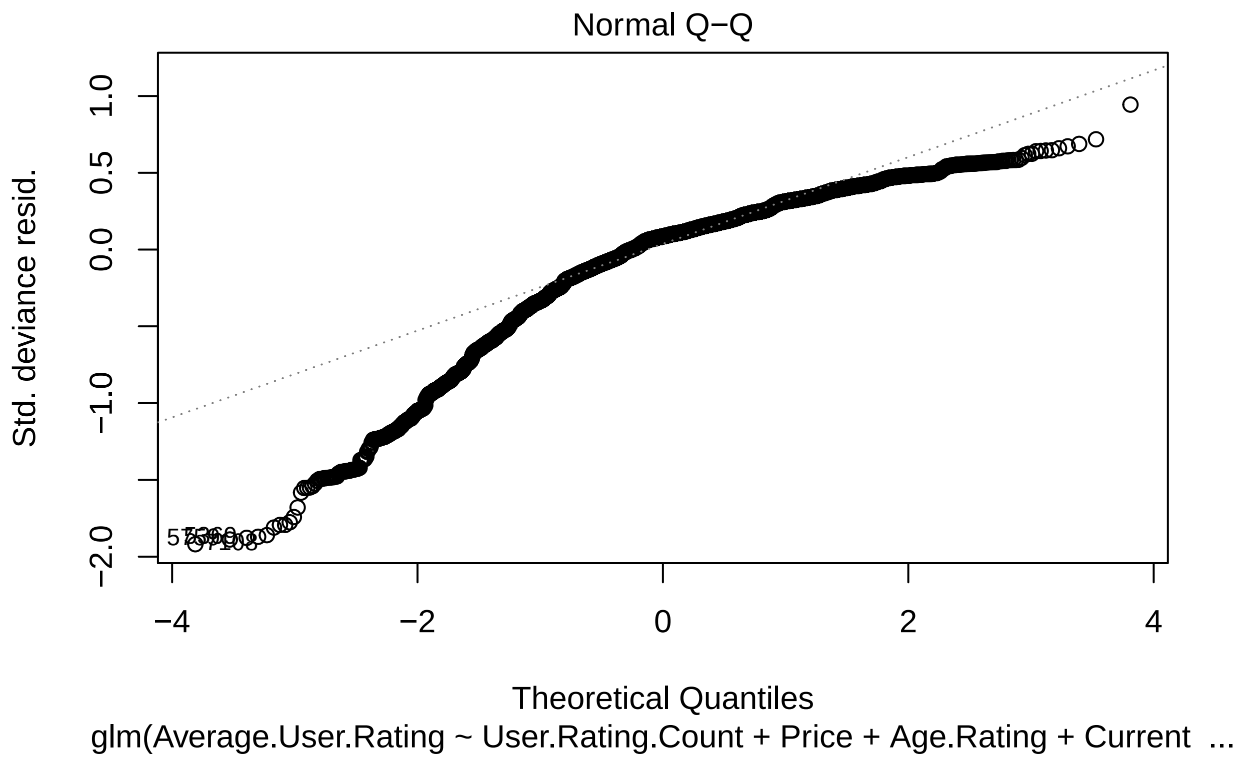
Modeling

Poisson Model

$$log(\frac{E(Rating)}{t}) = \alpha + \beta_1 * Count + \beta_2 * Price + \beta_3 * Age + \beta_4 * Current + \beta_5 * Original + \beta_6 * Word + \beta_7 * Lang + \beta_8 * Size$$

```
##
## Call:
## glm(formula = Average.User.Rating ~ User.Rating.Count + Price +
##     Age.Rating + Current + Original + word + lang + Size.MB,
##     family = poisson, data = app)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.91673  -0.15300   0.08621   0.22706   0.93217
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)        1.046e+00  2.633e-01   3.973 7.11e-05 ***
## User.Rating.Count  1.031e-07  1.186e-07   0.870   0.3844
## Price              4.670e-04  3.491e-03   0.134   0.8936
## Age.Rating17+     -4.141e-02  3.382e-02  -1.224   0.2208
## Age.Rating4+       6.566e-03  1.665e-02   0.394   0.6933
```

```
## Age.Rating9+       4.388e-03  1.915e-02   0.229   0.8187
## Current2009       -2.750e-02  3.108e-01  -0.088   0.9295
## Current2010        3.767e-02  2.924e-01   0.129   0.8975
## Current2011        7.341e-02  2.839e-01   0.259   0.7960
## Current2012        1.070e-01  2.804e-01   0.382   0.7027
## Current2013        9.641e-02  2.793e-01   0.345   0.7299
## Current2014        9.809e-02  2.788e-01   0.352   0.7250
## Current2015        1.341e-01  2.785e-01   0.482   0.6301
## Current2016        1.374e-01  2.782e-01   0.494   0.6214
## Current2017        1.759e-01  2.780e-01   0.633   0.5269
## Current2018        1.628e-01  2.781e-01   0.585   0.5584
## Current2019        1.982e-01  2.780e-01   0.713   0.4759
## Original2009       3.070e-03  9.971e-02   0.031   0.9754
## Original2010       6.133e-02  9.705e-02   0.632   0.5274
## Original2011       1.518e-01  9.428e-02   1.610   0.1074
## Original2012       1.819e-01  9.296e-02   1.957   0.0504 .
## Original2013       1.877e-01  9.234e-02   2.032   0.0421 *
## Original2014       1.966e-01  9.211e-02   2.134   0.0329 *
## Original2015       1.932e-01  9.214e-02   2.097   0.0360 *
## Original2016       1.897e-01  9.191e-02   2.064   0.0390 *
## Original2017       1.985e-01  9.199e-02   2.158   0.0309 *
## Original2018       1.963e-01  9.241e-02   2.124   0.0337 *
## Original2019       2.052e-01  9.395e-02   2.184   0.0290 *
## word               7.232e-05  3.786e-05   1.910   0.0561 .
## lang              -3.101e-04  9.082e-04  -0.341   0.7327
## Size.MB           -1.911e-05  2.646e-05  -0.722   0.4703
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1084.13  on 7219  degrees of freedom
## Residual deviance:  977.55  on 7189  degrees of freedom
## AIC: Inf
##
## Number of Fisher Scoring iterations: 4
```



The output of the poisson model shows that *Original* is statistically significant in this model. However, the QQ plot shows that most points do not fall on the line; all points form a curve and existing extreme value on both sides of the line. And the AIC is infinite. Therefore, this model is not fitted well.

Binomial Model

$$logit(p) = \alpha + \beta_1 * Count + \beta_2 * Price + \beta_3 * Age + \beta_4 * Current + \beta_5 * Original + \beta_6 * Word$$

```
##
## Call:
## glm(formula = as.factor(Rating) ~ User.Rating.Count + Age.Rating +
##     Current + Original + word, family = binomial, data = app)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.9985  -0.6145   0.6187   0.7675   1.9626
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -1.487e+01  3.942e+02  -0.038 0.969914
## User.Rating.Count  1.396e-05  3.515e-06   3.972 7.12e-05 ***
## Age.Rating17+     -5.265e-01  1.549e-01  -3.399 0.000677 ***
## Age.Rating4+       1.044e-02  8.312e-02   0.126 0.900067
## Age.Rating9+      -8.527e-02  9.711e-02  -0.878 0.379871
## Current2009       -1.279e-01  4.481e+02   0.000 0.999772
## Current2010        1.172e+01  3.942e+02   0.030 0.976276
## Current2011        1.215e+01  3.942e+02   0.031 0.975411
## Current2012        1.253e+01  3.942e+02   0.032 0.974646
## Current2013        1.260e+01  3.942e+02   0.032 0.974498
## Current2014        1.266e+01  3.942e+02   0.032 0.974377
```

```
## Current2015        1.307e+01  3.942e+02   0.033 0.973563
## Current2016        1.301e+01  3.942e+02   0.033 0.973666
## Current2017        1.344e+01  3.942e+02   0.034 0.972811
## Current2018        1.334e+01  3.942e+02   0.034 0.973004
## Current2019        1.386e+01  3.942e+02   0.035 0.971952
## Original2009       1.570e-01  4.501e-01   0.349 0.727201
## Original2010       1.259e+00  4.317e-01   2.917 0.003531 **
## Original2011       2.116e+00  4.254e-01   4.975 6.53e-07 ***
## Original2012       2.474e+00  4.219e-01   5.862 4.56e-09 ***
## Original2013       2.533e+00  4.189e-01   6.048 1.46e-09 ***
## Original2014       2.573e+00  4.182e-01   6.152 7.67e-10 ***
## Original2015       2.524e+00  4.179e-01   6.039 1.55e-09 ***
## Original2016       2.482e+00  4.166e-01   5.956 2.58e-09 ***
## Original2017       2.497e+00  4.174e-01   5.983 2.19e-09 ***
## Original2018       2.451e+00  4.200e-01   5.837 5.33e-09 ***
## Original2019       2.337e+00  4.320e-01   5.411 6.28e-08 ***
## word               1.485e-03  2.077e-04   7.146 8.93e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 8285.0  on 7219  degrees of freedom
## Residual deviance: 7529.2  on 7192  degrees of freedom
## AIC: 7585.2
##
## Number of Fisher Scoring iterations: 13
```

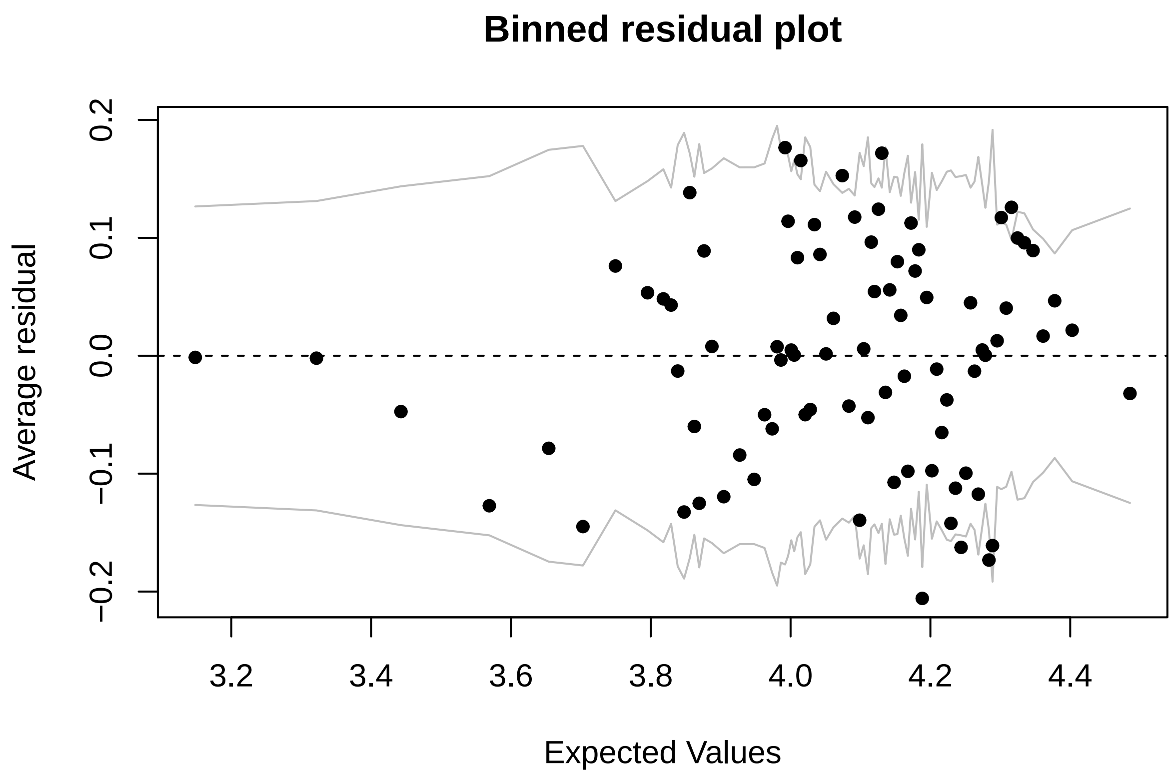
The binomial model uses the binary variable of rating; 1 means game rating is 4 or higher, and 0 means game rating is 3.5 or lower. Based on AIC, this model is selected by stepwise regression automatically. This model does not contain the *language* and *word*, which is the total number of descriptions of the game. This result is consistent with EDA.

Multilevel Linear Model with Random Intercept

$$Rating = \alpha_{(current+original)} + \beta_1 * Count + \beta_2 * Price + \beta_3 * Age + \beta_4 * Word + \beta_5 * Lang + \beta_6 * Size$$

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## Average.User.Rating ~ (1 | Current) + (1 | Original) + scale(User.Rating.Count) +
##     Price + Age.Rating + word + lang + Size.MB
##    Data: app
##
## REML criterion at convergence: 15599.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.6679 -0.4343  0.2523  0.6563  2.4652
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Current  (Intercept) 0.03398  0.1843
##  Original (Intercept) 0.08762  0.2960
##  Residual             0.49705  0.7050
## Number of obs: 7220, groups:  Current, 12; Original, 12
```

```
##
## Fixed effects:
##                            Estimate Std. Error t value
## (Intercept)               3.717e+00  1.062e-01  35.020
## scale(User.Rating.Count)  2.146e-02  8.392e-03   2.557
## Price                     2.065e-03  5.007e-03   0.413
## Age.Rating17+            -1.669e-01  4.759e-02  -3.507
## Age.Rating4+              2.579e-02  2.378e-02   1.085
## Age.Rating9+              1.711e-02  2.738e-02   0.625
## word                      3.023e-04  5.454e-05   5.543
## lang                     -1.190e-03  1.296e-03  -0.918
## Size.MB                  -7.646e-05  3.770e-05  -2.028
##
## Correlation of Fixed Effects:
##             (Intr) s(U.R. Price  A.R17+ Ag.R4+ Ag.R9+ word   lang
## scl(Us.R.C)  0.010
## Price       -0.008  0.030
## Age.Rtng17+ -0.086 -0.015  0.046
## Age.Ratng4+ -0.203 -0.003  0.040  0.363
## Age.Ratng9+ -0.149 -0.033  0.024  0.304  0.631
## word        -0.115 -0.024 -0.094  0.031  0.126 -0.011
## lang        -0.031 -0.056  0.050  0.017  0.023 -0.005 -0.056
## Size.MB     -0.051 -0.026 -0.225  0.031  0.200  0.091 -0.126 -0.029
```



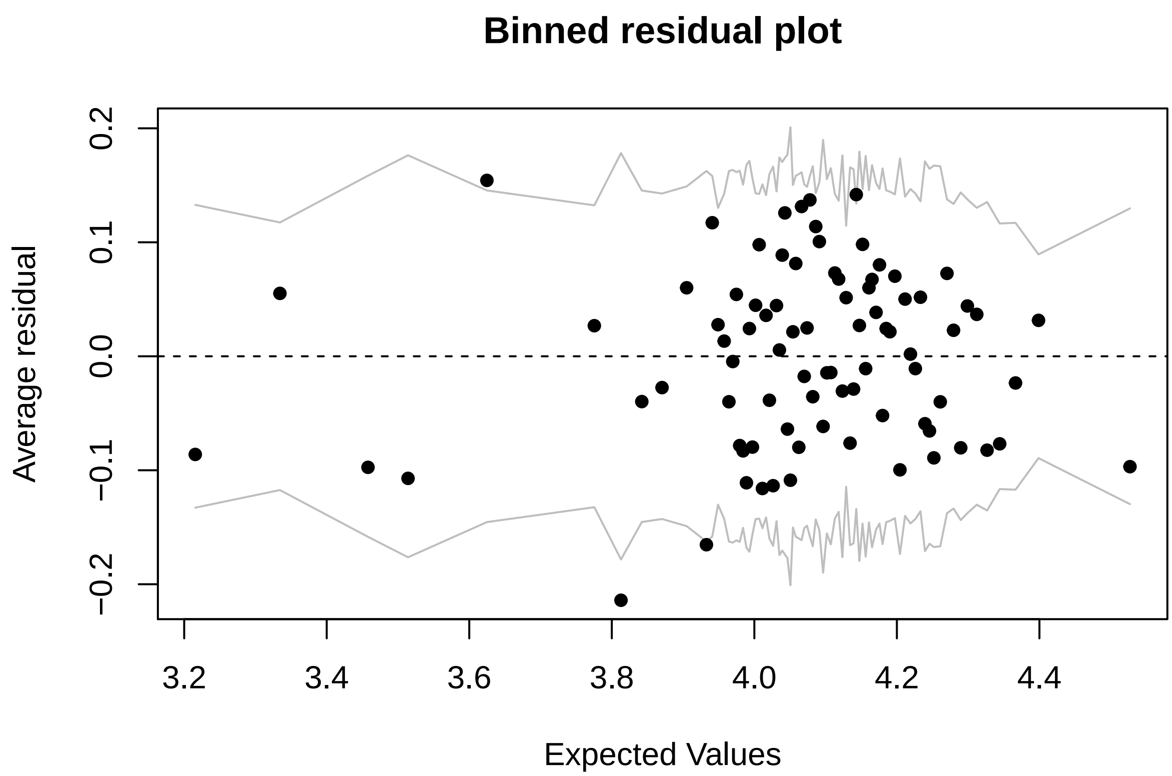
This model has added two random factors, *Original* and *Current*. From the estimates for the fixed effect, for one increase in the number of reviews, there is a 0.021 increase on game rating; for every one dollar increase in the price, there is a 0.002 increase on game rating, and games with age-rating 4+ have a greater chance of getting high scores. Binned residual plot shows some points fall outside the confidence bands, and the most points are placed at the right side.

Multilevel Linear Model with Random Slope and Intercept

$Rating = \alpha_{(original)} + \beta_1 * Count_{(current)} + \beta_2 * Price + \beta_3 * Age + \beta_4 * Word + \beta_5 * Lang + \beta_6 * Size + \mu_{current} * Count$

```
## boundary (singular) fit: see ?isSingular
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Average.User.Rating ~ (0 + User.Rating.Count | Current) + (1 |
##     Original) + scale(User.Rating.Count) + Price + Age.Rating +
##     word + lang + Size.MB
##    Data: app
##
## REML criterion at convergence: 15746.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.5729 -0.4392  0.2718  0.6412  2.1467
##
## Random effects:
##  Groups   Name              Variance  Std.Dev.
##  Current  User.Rating.Count 1.698e-10 1.303e-05
##  Original (Intercept)       1.209e-01 3.478e-01
##  Residual                   5.076e-01 7.125e-01
## Number of obs: 7220, groups:  Current, 12; Original, 12
##
## Fixed effects:
##                            Estimate Std. Error t value
## (Intercept)               3.830e+00  1.053e-01  36.352
## scale(User.Rating.Count)  2.336e-01  1.968e-01   1.187
## Price                     7.788e-03  5.008e-03   1.555
## Age.Rating17+            -1.729e-01  4.807e-02  -3.598
## Age.Rating4+              1.482e-02  2.399e-02   0.618
## Age.Rating9+              8.790e-04  2.760e-02   0.032
## word                      4.119e-04  5.413e-05   7.609
## lang                      7.903e-04  1.295e-03   0.610
## Size.MB                  -5.359e-05  3.796e-05  -1.412
##
## Correlation of Fixed Effects:
##             (Intr) s(U.R. Price  A.R17+ Ag.R4+ Ag.R9+ word   lang
## scl(Us.R.C)  0.135
## Price       -0.019  0.017
## Age.Rtng17+ -0.088  0.005  0.045
## Age.Ratng4+ -0.207  0.010  0.041  0.362
## Age.Ratng9+ -0.149  0.001  0.026  0.304  0.630
## word        -0.128 -0.009 -0.101  0.035  0.139 -0.001
## lang        -0.038  0.003  0.049  0.021  0.031  0.004 -0.085
## Size.MB     -0.053 -0.005 -0.225  0.033  0.205  0.096 -0.142 -0.042
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```



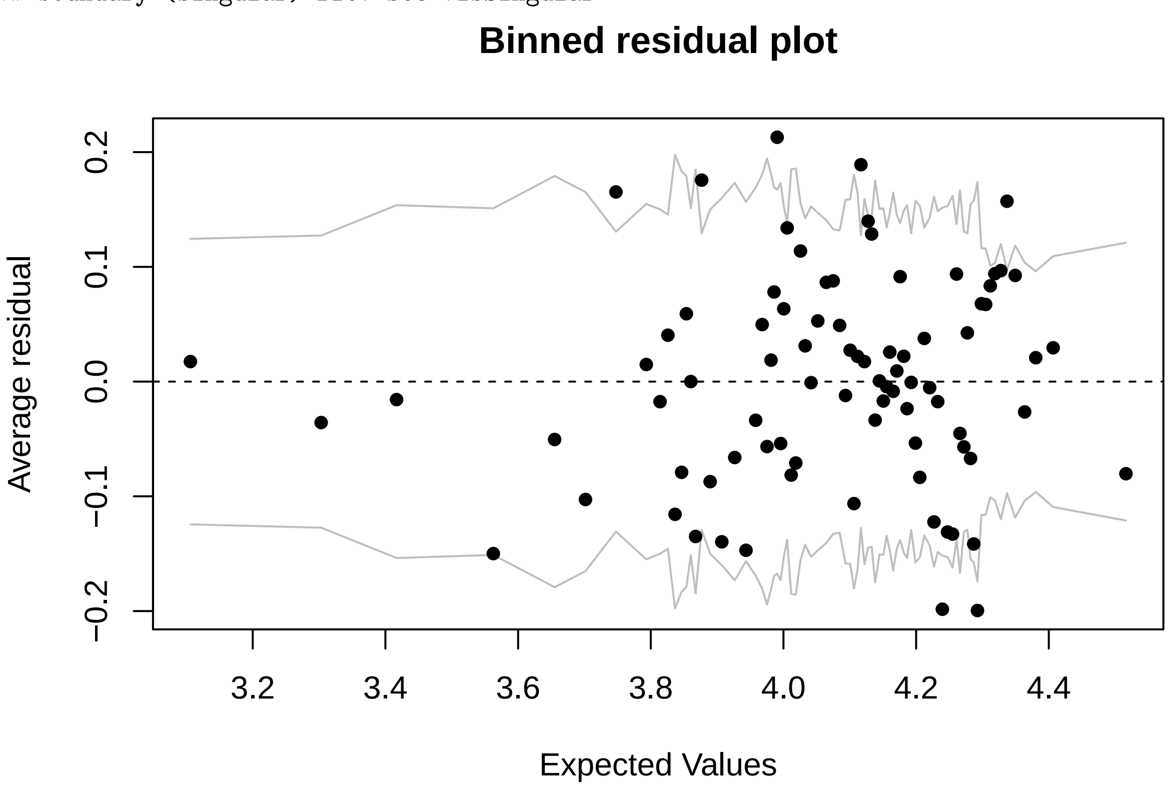
This model has added one random factor, $Original$, on intercept and one random slope for $Count$, which means that this model are allowing the slope of regression equation to vary by last release date of games. Since it is intuitive to assume that total number of reviews varies from year to year. From the estimates for the fixed effect, for one increase in the number of reviews, there is a 0.234 increase on game rating; for every one dollar increase in the price, there is a 0.008 increase on game rating, and games with age-rating 4+ have a greater chance of getting high scores. Binned residual plot shows only three points fall outside the confidence bands, but more points are placed at the right side.

$$Rating = \alpha_{(original)} + \beta_1 * Count + \beta_2 * Price + \beta_3 * Age + \beta_4 * Word + \beta_5 * Lang + \beta_6 * Size + \mu_{original} * Count + \mu_{current} * Count$$

```
## boundary (singular) fit: see ?isSingular
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Average.User.Rating ~ (1 + User.Rating.Count | Current) + (1 |
##     Original) + scale(User.Rating.Count) + Price + factor(Age.Rating) +
##     word + lang + Size.MB
##    Data: app
##
## REML criterion at convergence: 15611.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.6871 -0.4383  0.2504  0.6525  2.4420
##
## Random effects:
##  Groups   Name              Variance  Std.Dev.  Corr
##  Current  (Intercept)       2.543e-02 1.595e-01
##           User.Rating.Count 3.981e-09 6.309e-05 0.25
```

```
##  Original (Intercept)       1.469e-01 3.833e-01
##  Residual                   4.942e-01 7.030e-01
## Number of obs: 7220, groups:  Current, 12; Original, 12
##
## Fixed effects:
##                            Estimate Std. Error t value
## (Intercept)               3.733e+00  1.446e-01  25.817
## scale(User.Rating.Count)  2.455e-01  8.256e-01   0.297
## Price                     3.581e-03  4.997e-03   0.717
## factor(Age.Rating)17+    -1.625e-01  4.746e-02  -3.423
## factor(Age.Rating)4+      2.969e-02  2.374e-02   1.250
## factor(Age.Rating)9+      1.556e-02  2.732e-02   0.569
## word                      3.001e-04  5.443e-05   5.512
## lang                     -1.221e-03  1.293e-03  -0.944
## Size.MB                  -8.317e-05  3.762e-05  -2.211
##
## Correlation of Fixed Effects:
##             (Intr) s(U.R. Price  f(A.R)1 f(A.R)4 f(A.R)9 word   lang
## scl(Us.R.C)  0.515
## Price       -0.004  0.007
## fct(A.R)17+ -0.062  0.002  0.046
## fctr(A.R)4+ -0.147  0.005  0.042  0.363
## fctr(A.R)9+ -0.108  0.002  0.024  0.304   0.630
## word        -0.086 -0.006 -0.095  0.031   0.125  -0.012
## lang        -0.023  0.002  0.050  0.017   0.024  -0.005  -0.056
## Size.MB     -0.038 -0.003 -0.226  0.031   0.199   0.092  -0.126 -0.029
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```



This model has added two random factors, *Original* and *Current*, on intercept and kept *Current*

one random slope for *Count*. From the estimates for the fixed effect, for one increase in the number of reviews, there is a 0.246 increase on game rating; for every one dollar increase in the price, there is a 0.004 increase on game rating, and games with age-rating 4+ have a greater chance of getting high scores. Binned residual plot shows some points fall outside the confidence bands.

4. Result

Model Choice

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: app
## Models:
## mul1: Average.User.Rating ~ (1 | Current) + (1 | Original) + scale(User.Rating.Count) +
## mul1:     Price + Age.Rating + word + lang + Size.MB
## mul2: Average.User.Rating ~ (0 + User.Rating.Count | Current) + (1 |
## mul2:     Original) + scale(User.Rating.Count) + Price + Age.Rating +
## mul2:     word + lang + Size.MB
## mul3: Average.User.Rating ~ (1 + User.Rating.Count | Current) + (1 |
## mul3:     Original) + scale(User.Rating.Count) + Price + factor(Age.Rating) +
## mul3:     word + lang + Size.MB
##      Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## mul1 12 15541 15623 -7758.3    15517
## mul2 12 15685 15768 -7830.5    15661   0.00      0          1
## mul3 14 15530 15627 -7751.1    15502 158.85      2     <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

According to the output of the ANOVA test for three multilevel models, I find that multilevel model 3 with random intercept and slope of *Count* is the best one among the three multilevel models. Since the multilevel model 3 has the lowest AIC and its P-value is significant.

5. Discussion

Implication

In general, a high-score game will be downloaded by many people, so there will be many reviews. From the result of the multilevel model, the coefficient of *Count* shows that with one increase in the number of reviews, there is a 0.246 increase on game rating. Therefore, producers need to recommend players to write reviews.

The coefficient of *Price* indicates that expensive games are more likely to get a higher rating than the cheap ones. Therefore, I suggest that game producers can spend more time on making better games, which is consistent with common sense.

The size of the game and the number of languages available in the game will not impact the game rating a lot, and the game with a lower age limit may have a higher rating.

Limitation

The limitation of this model is that the dataset, which I used, only includes mobile strategy games and all of the data collected from the Apple store. Therefore, this model is not enough to predict the game rating for other types of games and games that download from different stores. Furthermore, this dataset does not include the total download of each game. Due to not everyone who download game will write a review, and the total number of rating is less than the total download, so this model is not very accurate.

Future Direction

For future works, I will try to include more data, which collected from other app stores and other types of games. And to find a more fitted model for this predicting problem.

6. Appendix

Fig.10 Most popular game developer

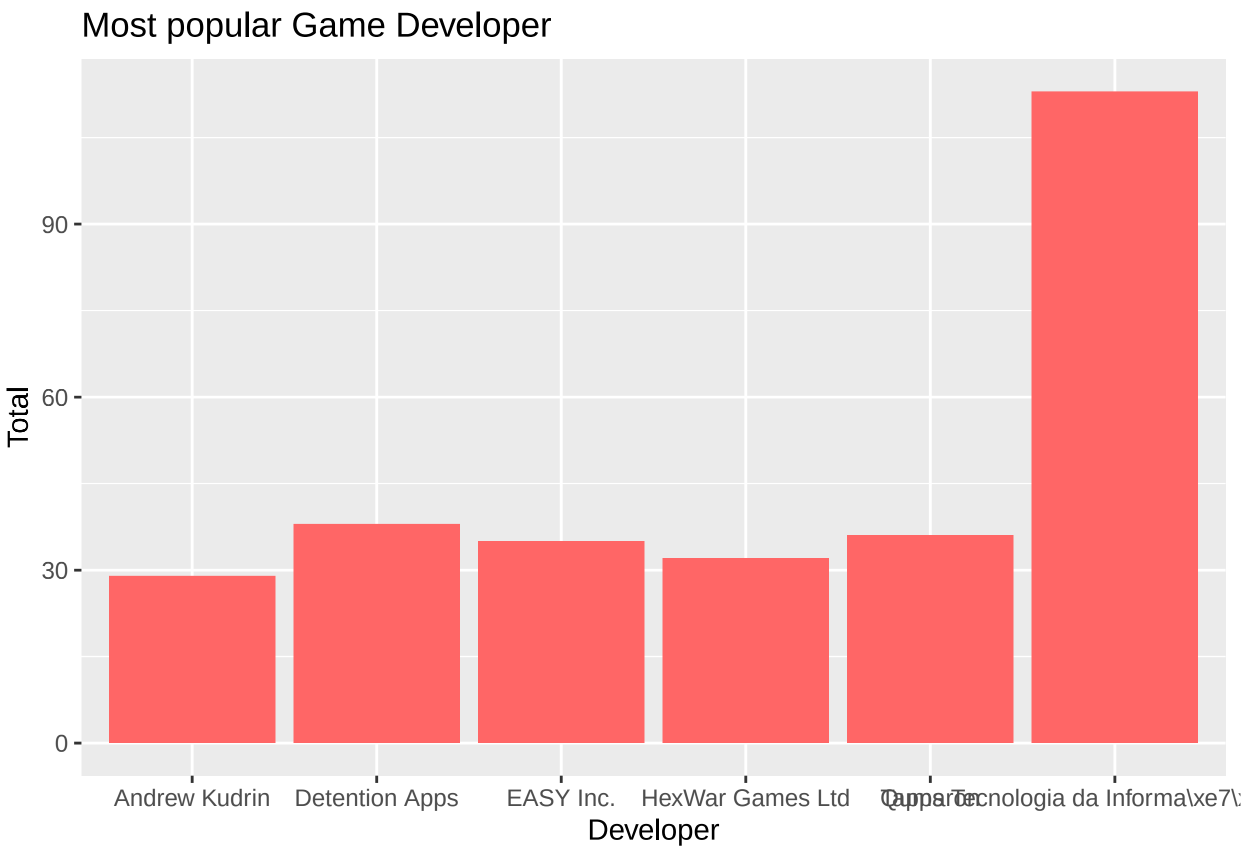
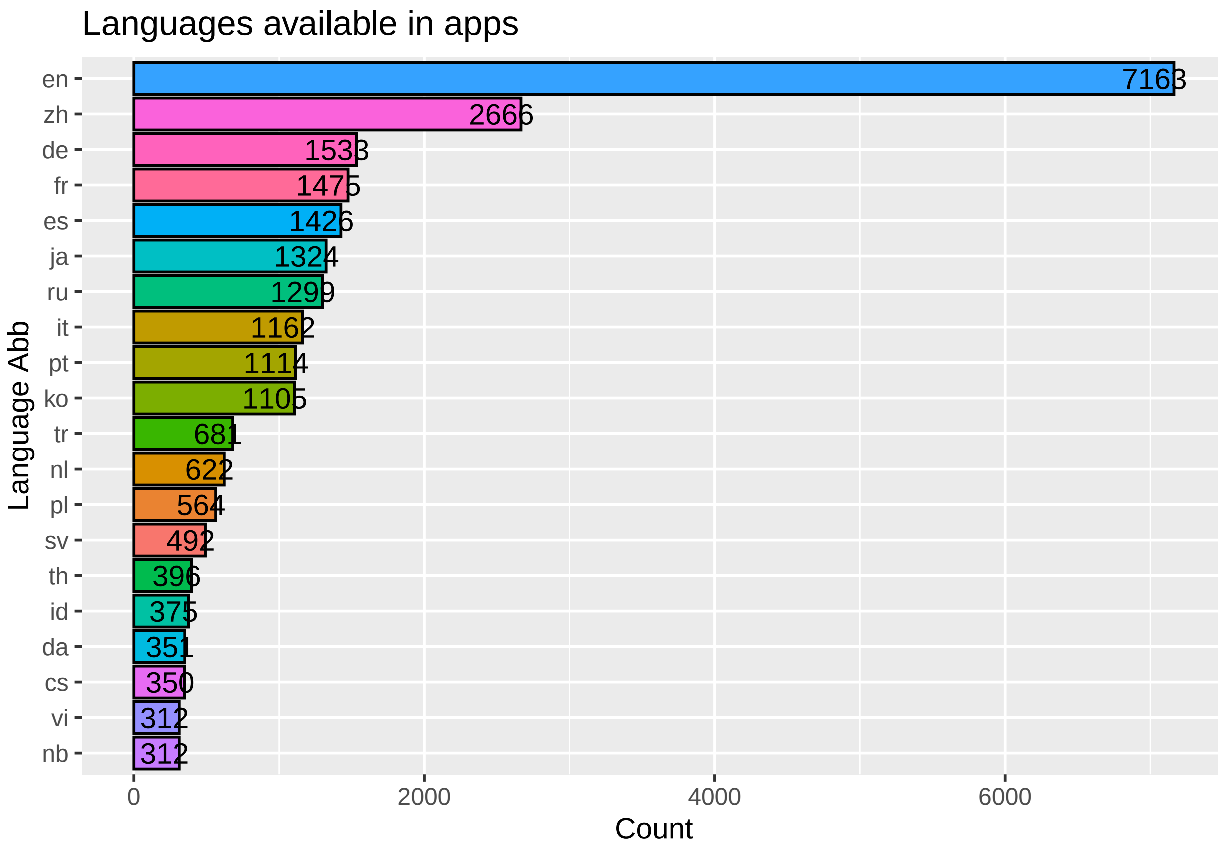


Fig.11 Total number of each available languages in games



7. Reference

https://www.kaggle.com/tristan581/17k-apple-app-store-strategy-games

https://en.wikipedia.org/wiki/Mobile_game

https://www.gamedesigning.org/gaming/great-games/